



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2015

Combined Label-Free Quantitative Proteomics and microRNA Expression Analysis of Breast Cancer Unravel Molecular Differences with Clinical Implications

Gámez-Pozo, Angelo ; Berges-Soria, Julia ; Arevalillo, Jorge M ; Nanni, Paolo ; López-Vacas, Rocío ;
Navarro, Hilario ; Grossmann, Jonas ; Castaneda, Carlos A ; Main, Paloma ; Díaz-Almirón, Mariana ;
Espinosa, Enrique ; Ciruelos, Eva ; Fresno Vara, Juan Ángel

Abstract: Better knowledge of the biology of breast cancer has allowed the use of new targeted therapies, leading to improved outcome. High-throughput technologies allow deepening into the molecular architecture of breast cancer, integrating different levels of information, which is important if it helps in making clinical decisions. microRNA (miRNA) and protein expression profiles were obtained from 71 estrogen receptor-positive (ER(+)) and 25 triple-negative breast cancer (TNBC) samples. RNA and proteins obtained from formalin-fixed, paraffin-embedded tumors were analyzed by RT-qPCR and LC/MS-MS, respectively. We applied probabilistic graphical models representing complex biologic systems as networks, confirming that ER(+) and TNBC subtypes are distinct biologic entities. The integration of miRNA and protein expression data unravels molecular processes that can be related to differences in the genesis and clinical evolution of these types of breast cancer. Our results confirm that TNBC has a unique metabolic profile that may be exploited for therapeutic intervention.

DOI: <https://doi.org/10.1158/0008-5472.CAN-14-1937>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-120456>

Journal Article

Accepted Version

Originally published at:

Gámez-Pozo, Angelo; Berges-Soria, Julia; Arevalillo, Jorge M; Nanni, Paolo; López-Vacas, Rocío; Navarro, Hilario; Grossmann, Jonas; Castaneda, Carlos A; Main, Paloma; Díaz-Almirón, Mariana; Espinosa, Enrique; Ciruelos, Eva; Fresno Vara, Juan Ángel (2015). Combined Label-Free Quantitative Proteomics and microRNA Expression Analysis of Breast Cancer Unravel Molecular Differences with Clinical Implications. *Cancer Research*, 75(11):2243-2253.

DOI: <https://doi.org/10.1158/0008-5472.CAN-14-1937>

Combined label-free quantitative proteomics and microRNA expression analysis of breast cancer unravel molecular differences with clinical implications

Angelo Gámez-Pozo¹, Julia Berges-Soria¹, Jorge M. Arevalillo², Paolo Nanni³, Rocío López-Vacas¹,
Hilario Navarro², Jonas Grossmann³, Carlos A. Castaneda⁴, Paloma Main⁵, Mariana Díaz-Almirón⁶,
Enrique Espinosa⁷, Eva Ciruelos⁸ and Juan Ángel Fresno Vara^{1,5}

¹Molecular Oncology & Pathology Lab, Instituto de Genética Médica y Molecular-INGEMM, Instituto de Investigación Hospital Universitario La Paz-IdiPAZ. Pº de la Castellana 261. 28046 Madrid. Spain.

²Department of Statistics, Operational Research and Numerical Analysis, University Nacional Educacion a Distancia (UNED). Paseo Senda del Rey 9, 28040 Madrid Spain.

³Functional Genomics Centre Zurich, University of Zurich/ETH Zurich, Winterthurerstrasse 190, CH-8057. Zurich, Switzerland.

⁴ Departamento de Investigación. Instituto Nacional de Enfermedades Neoplásicas, Av. Angamos Este 2520, Lima 34, Surquillo - Lima - Peru.

⁵Department of Statistics and Operations Research. Faculty of Mathematics. Complutense University of Madrid. Plaza de las Ciencias, 3, 28040. Madrid, Spain.

⁶Biostatistics Unit, Instituto de Investigación Hospital Universitario La Paz-IdiPAZ. Pº de la Castellana 261. 28046. Madrid. Spain.

⁷Medical Oncology Service, Instituto de Investigación Hospital Universitario La Paz-IdiPAZ. Pº de la Castellana 261. 28046. Madrid. Spain.

⁸Medical Oncology Service, Instituto de Investigación Hospital Universitario Doce de Octubre-i+12. Avda. de Córdoba, s/n. 28041Madrid. Spain.

[§]Corresponding author:

Dr. Juan Ángel Fresno Vara

juanangel.fresno@salud.madrid.org; Phone: +34 - 91 2071010 - Ext: 255; Fax: +34 - 91 2071040

Instituto de Genética Médica y Molecular (INGEMM)

Hospital Universitario La Paz-IdiPAZ

Paseo de la Castellana, 261-28046 - Madrid (Spain)

Running title:

Proteomics and miRNA expression analysis of breast cancer.

Keywords:

Breast cancer, FFPE, label-free proteomics, microRNA, systems biology, molecular medicine, therapeutic targets.

Financial support:

The authors would like to acknowledge funding from grants PS09/01597 and PI12/00444 from Instituto de Salud Carlos III, Spanish Economy and Competitiveness Ministry, Spain. This work has been also supported by the PRIME-XS project, grant agreement number 262067, funded by the European Union 7th Framework Programme. AG-P and RL-V are supported by Instituto de Salud Carlos III, Spanish Economy and Competitiveness Ministry grants CA12/00258 and CA12/00264 respectively. IdiPAZ and I+12 Biobanks are supported by Instituto de Salud Carlos III, Spanish Economy and Competitiveness Ministry (RD09/0076/00073 and RD09/0076/00118 respectively) and Farmaindustria, through the Cooperation Program in Clinical and Translational Research of the Community of Madrid.

Competing Interests:

JAFV, EE and AG-P are shareholders in Biomedica Molecular Medicine SL. The other authors declare that they have no competing interests.

Abstract

Better knowledge of the biology of breast cancer has allowed the use of new targeted therapies, leading to improved outcome. High-throughput technologies allow deepening into the molecular architecture of breast cancer, integrating different levels of information, which is important if it helps in making clinical decisions. MicroRNA and protein expression profiles were obtained from 71 estrogen receptor-positive and 25 triple-negative breast cancer samples. RNA and proteins obtained from formalin-fixed, paraffin-embedded tumors were analyzed by RT-qPCR and LC-MS/MS respectively. We applied probabilistic graphical models representing complex biological systems as networks, confirming that estrogen receptor-positive and triple-negative breast cancer subtypes are distinct biological entities. The integration of miRNA and protein expression data unravels molecular processes that can be related to differences in the genesis and clinical evolution of these types of breast cancer. Our results confirm that triple-negative breast cancer has a unique metabolic profile that may be exploited for therapeutic intervention.

Introduction

Breast cancer is a major health issue in developed countries. Early diagnosis and the use of adjuvant therapies have contributed to improve survival, but still over 95,000 women are expected to die of breast cancer in the European Union in 2015 (1). In the last decades, the death rate from breast cancer has decreased gradually, due to early diagnosis and to the availability of adjuvant chemotherapy and hormonal therapy. Early diagnosis allows the detection of *in situ* or very small infiltrating tumors, which have an excellent prognosis, whereas adjuvant therapy decreases the rate of relapse. For this reasons, factors determining the use of such therapies are key elements in patient management. Knowledge on the molecular biology of breast cancer has recently challenged the way in which oncologists make decisions on adjuvant chemotherapy. A number of techniques, such as gene profiling, proteomics and microRNA analysis have been used to study this disease.

Gene profiling in breast cancer has received increasing attention in recent years as a way to provide prognostic and predictive information. Since the description of a 70-gene profile in 2002 by van't Veer and colleagues, this and other gene signatures such as Recurrence Score have made their way into the clinic very rapidly, because they provide clinically useful information (2).

Proteomics would enable the unbiased comparison of different cellular states in biology and medicine at a systems-wide level (3). Proteome analysis heavily relies on mass spectrometry (MS). MS-based proteomics is starting to mature and to deliver through a combination of developments in instrumentation, sample preparation and computational analysis. These advances allow the identification of thousands of proteins from tissue amounts compatible with clinical routine, which is relevant for the study of complex diseases. For example, thousands of mutations have been described associated with cancer (4), but the exact relationship between genomic variation and

the resulting phenotype in each individual tumor remains unknown. Deep proteomics analyses are necessary to characterize the complete scenario of signaling pathways and biological processes altered as a result of the specific mutation set produced in each tumor.

MicroRNAs play an important role as regulators of gene expression, controlling many biological processes such as growth, development, differentiation and apoptosis. This has led to considerable interest in the development of miRNA expression profiles as a new class of biomarkers in cancer, cardiovascular and autoimmune diseases. Measuring miRNA expression can also be useful for systems-level studies of gene regulation, especially when miRNA measurements are combined with mRNA profiling and other genome-scale data (5).

A major limitation for the widespread use of high-throughput techniques is tissue availability. Fresh or fresh-frozen specimens are ideal for molecular profiling but rarely available in the clinic. In recent years, we and others have demonstrated that formalin-fixed, paraffin embedded (FFPE) tissue can be readily analyzable by MS-based technologies (6, 7). On the other hand, previous studies demonstrate the possibility of performing experiments to analyze the expression of miRNAs (8) in FFPE tissues. FFPE samples also constitute the basis for routine workup in pathology departments all over the world.

In this paper, we combined quantitative proteomics with miRNA expression analyses in a series of breast cancer samples with appropriate clinical information. We demonstrate that it is possible to perform differential protein expression analyses by LC-MS/MS on tens of FFPE tumors. Protein patterns confirm that estrogen receptor-positive (ER+) and triple-negative (TNBC) breast cancer subtypes are distinct biological entities (9). Also, we were able to profile miRNA expression in the same series of samples, which led us to conduct a probabilistic graphical models analysis to

integrate miRNA and protein expression data. This systems-level study sheds light on the molecular processes differentially regulated in these two main subtypes of breast cancer.

Materials & Methods

Study design and sample description

Our primary objectives were to develop a method for the study of both proteins and microRNA expression pattern from a systems biology point of view, and to identify differences in biological processes between the two main groups of breast cancer patients: estrogen receptor-positive, ER+, and triple-negative breast cancer, TNBC. One hundred and six FFPE samples from patients diagnosed of breast cancer were retrieved from I+12 Biobank (RD09/0076/00118) and from IdiPAZ Biobank (RD09/0076/00073), both integrated in the Spanish Hospital Biobank Network (RetBioH; www.redbiobancos.es). The histopathological features of each sample were reviewed by an experienced pathologist to confirm diagnosis and tumor content. Eligible samples had to include at least 50% of tumor cells. Approval from the Ethical Committees of Hospital Doce de Octubre and Hospital Universitario La Paz was obtained for the conduct of the study.

Total protein preparation

Proteins were extracted from FFPE samples as previously described (10). Briefly, FFPE sections were deparaffinized in xylene and washed twice with absolute ethanol. Protein extracts from FFPE samples were prepared in 2% SDS buffer using a protocol based on heat-induced antigen retrieval (6). Protein concentration was determined using the MicroBCA Protein Assay Kit (Pierce-Thermo Scientific). Protein extracts (10 µg) were digested with trypsin (1:50) and SDS was removed from digested lysates using Detergent Removal Spin Columns (Pierce). Peptide samples were further desalted using ZipTips (Millipore), dried, and resolubilized in 15 µL of a 0.1% formic acid and 3% acetonitrile solution before MS analysis.

Liquid chromatography - mass spectrometry analysis

Samples (4 μ L) were analyzed on a LTQ-Orbitrap Velos hybrid mass spectrometer (Thermo Fischer Scientific, Bremen, Germany) coupled to NanoLC-Ultra system (Eksigent Technologies, Dublin, CA, USA). Peptide separation was performed on a self-made column (75 μ m \times 150 mm) packed with Magic RP C18 AQ, 200A, 3 μ m beads (Bischoff GmbH, Leonberg, Germany) at a flow rate of 250 nL/min. Solvent composition was 0.1% formic acid for channel A, and 0.1% formic acid and 99.9% acetonitrile for channel B. The peptides were eluted with a gradient of 5 to 30% acetonitrile in 95 minutes. The mass spectrometer was operated in data-dependent mode (DDA), acquiring a full-scan MS spectra (300–1700 m/z) at a resolution of 30,000 at 400 m/z after accumulation to a target value of 1,000,000, followed by CID (collision-induced dissociation) fragmentation on the twenty most intense signals per cycle. CID spectra were acquired using normalized collision energy of 35 and a maximum injection time of 50 ms. The automatic gain control (AGC) was set to 10,000 ions. Charge state screening was enabled and singly and unassigned charge states were rejected. Precursor masses previously selected for MS/MS measurement were excluded from further selection for 45 s, and the exclusion window was set at 10 ppm. The size of the exclusion list was set to a maximum of 500 entries. The samples were acquired using internal lock mass calibration on m/z 429.088735 and 445.120025.

Label-free protein expression data processing

The acquired raw MS data were processed by MaxQuant (version 1.2.7.4), followed by protein identification using the integrated Andromeda search engine. Spectra were searched against a UniProtKB/Swiss-Prot database for human (73,849 entries), including a set of and 260 common

protein contaminants (NCBI taxonomy ID 9606, download date 2011-12-13). Reversing the protein sequences was chosen as decoy option in MaxQuant. Carbamidomethylation of cysteine was set as fixed modification, while oxidation (M), deamidation (N, Q) and N-terminal protein acetylation were set as variable. Enzyme specificity was set to trypsin/P allowing a minimal peptide length of 7 amino acids and a maximum of two missed-cleavages. Precursor tolerance was set to 20 ppm, while fragment was set to 0.5 Da. The maximum false discovery rate (FDR) was set to 0.01 for peptides and 0.05 for proteins.

Label free quantification was performed setting a 2 minutes window for match between runs. The protein abundance was calculated on the basis of the normalized spectral protein intensity (LFQ intensity). Quantifiable proteins were defined as those detected in at least 75% of samples in at least one type of sample (either ER+ or TNBC samples) showing two or more unique peptides. Only quantifiable proteins were considered for subsequent analyses. Protein expression data were log2 transformed and missing values were replaced using data imputation for label-free data using default values. Finally, protein expression values were z-score transformed. Batch effects were estimated and corrected using ComBat (11).

All the mass spectrometry raw data files acquired in this study may be downloaded from Chorus (<http://chorusproject.org>) under the project name *Breast Cancer Proteomics*. The *peptides* output file from the MaxQuant analysis is provided as supplementary material.

RNA extraction and MicroRNA expression

Selected FFPE tumor specimens were cut into serial sections with a thickness of 10 µm. Total RNA was then isolated using the RecoverAll Total Nucleic Acid Isolation Kit (Life Technologies). Purified

RNA quality control for quantity and purity was assessed using an ND-1000 NanoDrop spectrophotometer (Thermo Fisher Scientific). A preliminary analysis was performed to compare miRNA expression measurements between paired FF and FFPE samples, as described previously (12). MicroRNA expression profiling was performed using a custom TaqMan Array MicroRNA Card (Life Technologies) containing 95 FFPE-reliable assays, including four housekeeping miRNAs identified used NorMean (12), plus one mandatory control. RNA concentration was adjusted to 166.7 ng/μL. For miRNA cDNA synthesis, 500 ng of total RNA was reverse-transcribed using the TaqMan MicroRNA Reverse Transcription Kit combined with the stem-loop Megaplex Primer Pools according to the manufacturer's protocol (Life Technologies). All PCR reactions were performed on the ABI Prism 7900HT Sequence detection system (Life Technologies). Average cycling threshold (Cq) values were obtained using SDS 2.2 software (Applied Biosystems) employing automatic baseline settings and a threshold of 0.2. The maximum Cq value was set at 40. Cq values were normalized using two out of four reference miRNAs (hsa-let-7d and hsa-let-7g) selected as previously described (12). One microRNA was excluded from subsequent analyses because no detection was found in most of the samples. The relative expression level of each target gene was expressed as $\Delta Cq = Cq_{ref} - Cq_{gene}$. Cq_{ref} was obtained calculating the geometric mean of two housekeeping miRs (let-7d and let-7g). Reference-normalized expression measurements were adjusted by defining the lowest expression value as 0, with subsequent 1-unit increases reflecting an approximate doubling of the RNA. Finally, values were z-scored.

Assessing molecular differences between ER+ and TNBC samples

Significance Analysis of Microarrays (SAM) was carried out to find differentially expressed proteins and miRNAs between ER+ and TNBC samples with a FDR below 5%. Hierarchical clusters were

constructed with the differentially expressed proteins or miRNAs between ER+ and TNBC samples identified by SAM, using Pearson Correlation and the average-linkage method. Then, we used miRWalk (15th March 2013 version) (13) to find validated targets of these miRNAs among the proteins differentially expressed between ER+ and TNBC samples.

Network construction

Statistical analyses were conducted to associate miRNAs and proteins. As a first approach to describe associations present in our data base, we choose probabilistic graphical models compatible with high-dimensionality. The result is an undirected graphical model with local minimum Bayesian Information Criterion (BIC) (14) obtained after executing the next steps: first the spanning tree with maximum likelihood is found and then (I), a forward search is performed by successively adding edges that reduce the BIC and still preserve the decomposability (15) of the initial graph (II). In the first stage, which learns a Markov tree structure from a random sample of a supposed multidimensional normal population, we use the extension of the Chow-Liu solution (16), according to which, for categorical data, the maximum likelihood structure is given by the maximum weight spanning tree (17) with empirical mutual information quantities (18) as edge weights. In the Gaussian case, a similar reasoning applies, but now the mutual information value is $-(1/2)\log(1-r^2)$, where r is the empirical correlation coefficient between the two variables (nodes) joined by the edge. Given that the algorithm is invariant under monotone transformations of the variables, r^2 can be used as a weight. In the second phase, we introduce the BIC criterion which penalizes more complex models and then, simpler graphs are generated. This is a fundamental objective in high-dimensional problems. Both methods are implemented in the open-source

statistical programming language R (19). In particular, the functions *minForest* and *stepw*, in the *gRapHD* package (20), are used for phases I and II, respectively.

Gene Ontology analyses

Protein-to-gene ID conversion was performed using Uniprot (<http://www.uniprot.org>) and DAVID. Gene ontology analyses were performed using the functional annotation chart tool provided by DAVID. We used “*homo sapiens*” as background list and selected only GOTERM-FAT gene ontology categories and Biocarta, KEGG and Panther pathways.

Functional node identification and activity measurement

To identify functional nodes within the probabilistic graphical models, we split it in several branches. Then, we used gene ontology analyses to investigate which function or functions were overrepresented in each branch. To measure the functional activity of each node, we calculated the mean expression of all the proteins included in one branch related with a concrete function. Differences in node activity between ER+ and TNBC samples were assessed by class comparison analyses.

Orthogonal verification

We used published array expression data of 1,296 primary breast carcinomas from two previously published works (21, 22). Batch effects between data sets were estimated and corrected using ComBat (11). After protein-to-gene ID conversion, all probes in dataset for each gene were

retrieved. Probes with higher coefficient of variation were selected when multiple probes were found for a single gene. Differences in node activity between ER+ and TNBC samples were assessed using a two-sample T-test analysis comparing expression. A Multivariate Permutations test was computed based on 1000 random permutations, allowing a confidence level of false discovery rate assessment of 90 % and a maximum allowed proportion of false-positive genes of 0.05

Cell culture

The ER+ cell lines MCF7 and T47D and the TNBC cell lines MDA-MB-231 and MDA-MB-468 were cultured in RPMI-1640 medium with phenol red (Gibco) supplemented with 10% heat-inactivated fetal bovine serum, 100 mg/mL penicillin and 100 mg/mL streptomycin. The TNBC cell line BT-20 was cultured in Minimum Essential Medium supplemented with 10% heat-inactivated fetal bovine serum, 100 mg/mL penicillin and 100 mg/mL streptomycin. All cell lines were cultured at 37 °C in a humidified atmosphere with 5% (v/v) CO₂ in air. MCF7, T47D BT-20 and MDA-MB-231 cell lines were kindly provided by Dra. Nuria Vilaboa (Hospital Universitario La Paz, previously obtained from ATCC in January 2014). MDA-MB-468 cell line was obtained from ATCC (July 2014). Cell lines were routinely monitored in our laboratory and authenticated by morphology and growth characteristics, tested for Mycoplasma and frozen, and passaged for fewer than 6 months before experiments.

Glucose and lactate measurements

250.000 cells were plated in 4 ml of fresh medium, and cultivated for 72h. Afterwards, supernatants were centrifuged to eliminate cellular debris. Lactate and glucose concentrations in cell supernatants and fresh media were measured on an amperometric electrode using the enzymes lactate oxidase and glucose oxidase respectively, in an ABL90 FLEX gas analyzer (Radiometer) with integrated gas, electrolyte, metabolite and CO-oximetry measurements. At least measurements from three different experiments are provided. Glucose consumption is the difference between glucose concentration in fresh medium and glucose concentration in cell supernatant after 72h of cell culture. Lactate production is the difference between lactate concentration in cell supernatant after 72h of cell culture and lactate concentration in fresh medium. Differences were assessed by Kruskal-Wallis test.

Statistical analyses and software suites

BRB-ArrayTools, SPSS v16 software package, GraphPad Prism 5.1 and R v2.15.2 (with the *Design* software package 0.2.3) were used for all statistical analyses. All p-values were two-sided and $p < 0.05$ was considered statistically significant. Expression data and network analyses were performed in MeV and Cytoscape software suites. BRB-ArrayTools has been developed by Dr. Richard Simon and BRB-ArrayTools Development Team.

Results

Study goal and patient's characteristics

Our study aims to develop a method to study both proteins and microRNA expression pattern from a systems biology point of view, to discover differences in biological processes between the two main groups of breast cancer patients: estrogen receptor-positive, ER+, and triple-negative breast cancer, TNBC. One hundred and six patients were included in the study. Clinical characteristics from these patients are provided in table 1. All patients were node positive; all of them were negative when tested for Her2 amplification using immunohistochemistry and fluorescent *in situ* hybridization when needed; and all of them received adjuvant chemotherapy, (either anthracycline-based or not) plus hormonal therapy in the ER+ group.

Protein extraction and MS analysis

We have previously demonstrated that our workflow is high reproducible (10). One hundred and two FFPE samples yielded enough protein to perform the MS analyses (typically, more than 0.5 milligrams of total protein was recovered from each sample). After MS workflow, 96 samples gave useful protein expression data (25 TNBCs and 71 ER+ tumors). A total of 3,239 protein groups were identified using Andromeda (Supplementary table 1), of which 1,095 presented at least two unique peptides and detectable expression in at least 75% of the samples in at least one type of sample (either ER+ or TNBC samples) (Supplementary table 2). No one decoy protein passed through these additional filters. Ten samples were excluded from the study; four of them did not yield enough protein to perform the MS analyses, and six did not reach the “mean minus twice the standard deviation”-threshold in the number of unique peptides identified.

Label-free proteomics analysis

Label-free quantification was performed using MaxQuant (Supplementary table 1). Expression values for each protein were analyzed using Significance Analysis of Microarrays (SAM). We found 100 protein differentially expressed between ER+ and TNBC samples with a FDR<5% (Supplementary table 3). Then, we built a hierarchical cluster using the expression values of these 100 proteins (Figure 1). Gene ontology analyses from these 100 proteins showed a relevant enrichment in extracellular matrix proteins (Supplementary table 4).

Orthogonal verification of differences between ER+ and TNBC samples

We used previously published gene expression data as verification of the results obtained using SAM (21, 22). After protein-to-gene ID conversion, 90 out of 100 proteins could be matched by this method. We found that the differences identified at the protein level were confirmed for 74 genes, which showed a significantly different expression in ER+ and ER- samples with a FDR<5%. A group of TNBC samples could not be defined in this verification series because HER2 status was not reported for most samples.

MicroRNA expression analysis

We analyzed the expression of 90 miRNAs in samples from the study with enough available tissue (Supplementary table 5 & 6). Expression values for each miRNA were analyzed using SAM. We found 19 miRNAs differentially expressed between ER+ and TNBC samples with an FDR<5%

(Supplementary table 7). miR-20a, miR-19a, miR-106a, miR-18a and miR-135b expression were higher in TNBC samples, whereas miR-190b, miR-375, miR-449a, miR-139-5p, miR-205, miR-342, miR-149, miR-625, miR-193b, miR-214, miR-125b, miR-199a-3p, miR-30a* and miR-30e* were more expressed in ER+ samples. Then, we built a hierarchical cluster using the expression values of these 19 miRNAs (Figure 2). As can be observed, TNBC samples cluster together.

Systems Biology of breast cancers

Both label-free protein quantification and microRNA expression data were available for 79 tumors (16 TNBC and 63 ER+ tumors). Once we identified proteins and miRNAs differentially expressed between ER+ and TNBC samples, we checked if any of these proteins had been previously described as a regulated target of any of these miRNAs using the validated targets module of miRWalk. We found that only five proteins were previously identified as targets of six miRNAs. This lack of information prompted us to explore new ways to generate relations, if not between miRNAs and proteins, between miRNAs and biological processes and pathways.

Protein and microRNA expression data from all samples were used in the probabilistic graphical models analyses, with no other *a priori* information. Then, the resulting graph was processed (Figure 3) looking for a functional structure, i.e., if the proteins and miRNAs included in each branch of the tree had some relationship regarding their function. Thus, we divided our graph in twelve branches and a core, and performed gene ontology analyses. We isolate branches based on the structure of the probabilistic graphical model starting from the outside, with a minimum of 40 proteins included in each branch to allow gene ontology analyses in each branch and the core independently. The structure of the probabilistic graphical model had a strong biological function basis, as ten out of twelve branches showed a significant enrichment in proteins related with one

or two specific biological functions (Figure 3). Gene ontology analyses also showed that the core includes proteins related with most of the categories identified in the branches, but no category was over-represented.

The next step was to identify which functions are differentially represented between ER+ and TNBC tumors. To do this, we established the level of expression of each function node using the mean expression of all the proteins included in a given branch that belong to a single functional group, building twelve protein functional nodes (Figure 3 and Supplementary Table 8). Then we performed class comparison analyses to assess which functional nodes are differentially activated between ER+ and TNBC samples. Seven nodes showed significant differences between both groups. Metabolism B, Proliferation, Protein Synthesis and Mitochondria B functional nodes showed increased activity in TNBC samples, whereas Mitochondria A, Metabolism A and Extracellular Matrix nodes showed increased activity in ER+ tumors (Figure 4 and Supplementary Table 8).

Then, we compared the proteins identified as differential between SAM and the probabilistic graphical model functional approach. Only twenty five proteins from the SAM analysis were also in the 239 proteins from the seven functional nodes differentially expressed between ER+ and TNBCs. Ultimately, we performed an orthogonal verification of the activity of these functional nodes in our validation dataset. Six out of seven nodes showed a significant difference of activation between ER+ and ER- samples.

Regarding the miRNAs, our first observation was that most miRNAs were included in two big miRNA groups, meaning that relations at the expression level are stronger between these microRNAs themselves than the relation between microRNAs and proteins. One of these groups contains every measured miRNA from the human mir-17-92 cluster, whereas the other contains

more than 50% of all the miRNAs analyzed in this study (Figure 3). The most outstanding observation was that fifteen miRNAs were included in branches composed mainly by proteins (Supplementary Table 8).

Function: Proliferation

We found two different functions mixed in the same branch: Proliferation and Extracellular Matrix (ECM), and two miRNAs embedded within: miR-301a and miR-331-5p. Thirteen proteins related with regulation of cell proliferation belong to the proliferation node. STAT1 is more expressed in TNBC samples than in ER+ ones, whereas PURA expression is decreased in TNBC samples accordingly with SAM analyses. Measurement of the functional activity of this node showed a higher proliferation activity in our TNBC sample series ($p < 0.01$), as well as in the validation cohort ($p < 0.001$).

Function: Extracellular matrix

We also found 39 proteins related with ECM and focal adhesion. Eight of the components of this functional node were also found as differential between ER+ and TNBC samples in SAM analysis. All of them (Decorin, VCAN, Mimecan, Biglycan, COMP, Lumican, Prolargin and Asporin) showed lower expression in TNBC samples. Thus, it is not surprising that the ECM functional node activity highly differs between ER+ and TNBC samples, both in the proteomics analysis ($p < 0.01$) and in the orthogonal validation ($p < 0.05$).

Function: Metabolism

We found two different branches related with metabolism, thus we evaluated these two functional metabolism nodes independently. Metabolism A functional node included 33 proteins. Among them, seven were identified as differentially expressed by SAM analyses. PHGDH, LDHB and P5CS (encoded by the ALDH18A1 gene) are more expressed in TNBC samples than in ER+, whereas DHPR, FBP1 and MCCB are decreased in TNBC samples. The functional activity of this node was higher in ER+ samples ($p < 0.01$) and was also significant in the orthogonal validation ($p < 0.001$). Three microRNAs were included in this functional node: miR-135b, miR-190b and miR-224. Differential expression analysis showed that miR-135b expression was higher in TNBC samples, whereas miR-190b was less expressed in TNBC samples. Metabolism A includes proteins related with glutamine metabolism, such as P5CS, IDH1 and GLUD1. GLUD1 protein expression is slightly lower in TNBC tumors (fold change=0.66, $p < 0.05$).

Metabolism B functional node comprised thirteen proteins, including GAPDH, PGK1, HSP90, LDHA and Pyruvate kinase among others. When we measured the functional activity of this node, it appeared increased in TNBC both in our proteomics data ($p < 0.01$) as in the orthogonal validation ($p < 0.001$). However, only one protein (UPGA) was included among the 100 proteins identified as differentially expressed between ER+ and TNBC. Only one microRNA (miR-449a) was included in this functional node, and it was also detected by SAM as down-regulated in TNBC samples. So, we identified a functional node with increased activity and a miRNA related to it with decreased expression (Figure 4), showing a negative correlation ($r = -0.45$, $p < 0.001$, $r^2 = 0.20$; Supplementary figure 1). Surprisingly, miR-449a expression was not significantly correlated to any of the thirteen proteins included in this functional node.

Function: Mitochondrial

We found two different branches related with mitochondrial activity, thus we evaluated these functional mitochondrial nodes independently. Mitochondria A functional node included 37 proteins. Among them, four were identified as differentially expressed by SAM analyses. CYBB is more expressed in TNBC than in ER+, whereas IVD, DCXR and HSD17B8 expression is decreased in TNBC samples. The functional activity of this node was higher in our ER+ samples ($p < 0.05$), and was also confirmed by the orthogonal validation ($p < 0.001$). Two microRNAs (miR-23b and miR-375) were included in this functional node. miR-375 expression was also found increased in ER+ vs. TNBCs in SAM analysis. Mitochondria B functional node comprises 33 proteins, including, among others, Citrate synthase, SDHA and DLST, all related with the TCA cycle. Its activity is increased in TNBC samples in both our proteomics data ($p < 0.05$) and in the orthogonal validation ($p < 0.001$). None of the studied microRNAs were included in this functional node. Two proteins from this node showed an increased expression in TNBC samples in SAM analysis: ATP6V1D and SHMT2.

Metabolic analyses in breast cancer cell lines

We measured both the glucose consumption and lactate production in five breast cancer cell lines: two ER+ cell lines (MCF7 and T47D) and three TNBC cell lines (MDA-MB-231, BT-20 and MDA-MB-468). TNBC cell lines showed higher glucose consumption than ER+ cell lines (Figure 5A). Lactate production was also higher in TNBC cell lines compared with ER+ ones (Figure 5B).

Discussion

In this work we demonstrate that it is technically feasible to analyze up to a hundred of clinical samples by high-throughput proteomics, which could help to place mass spectrometry at the same level than gene microarrays, or massively parallel sequencing technologies, in its capability to analyze enough samples and features to generate clinically useful molecular profiles. Ongoing studies are trying to complete the molecular characterization of breast cancers (23). One key aspect of these studies is that they will analyze hundreds of samples to extract meaningful information, both from the medical and biological point of view.

Our proteomics pipeline allowed us to detect 3,239 protein groups in FFPE breast tissues. We then applied stringent filters to avoid false detections and used only 1,095 proteins in subsequent analyses. It is remarkable that our label-free quantification approach can detect clinically relevant proteins such as HER2, STAT1 or PARP1. However, other clinically important markers such as ER, PR or Ki67, were not detected using this technique. One obvious limitation of proteomics when compared to genomics is that genomics can measure the expression of all known genes in the same experiment, whereas proteomics only provides a measurement of peptides that are both detected and identified. This is directly related to the wide dynamic range of protein expression which means that the amount of some proteins can be orders of magnitude higher than others. Unfortunately, ER and PR are among these proteins showing low relative abundance in cells. We compared the abundance of ESR1, PRG and MKI67 with some proteins identified in our MS experiments using available data from PaxDB (24). ESR1 and PGR abundance is ten times lower than the less abundant protein identified in our MS experiments. Ki67 abundance is comparable to other detected proteins. We did not detect/identify any single peptide from this protein in our MS runs, thus with our experimental conditions, Ki67 cannot be detected in FFPE breast cancer

samples. These limitations could be overcome soon, as the latest developments in the field are promising.

Label-free proteomics analysis

SAM analysis identified one hundred proteins differentially expressed between ER+ and TNBC samples. Some of these proteins with higher expression in TNBC samples have been previously described as over-expressed in these tumors, such as PHGDH (25), LDHB (26), S100A8, S100A9 (27) and GRP78 (28), whereas AGR2, AGR3 (29), PIP (30) and GPD1 (31), have been shown to be up-regulated in ER+ breast cancer tumors in concordance with our data. These results are strained by sample size. For this reason proteomic-based quantitation of potential biomarkers requires further validation using orthogonal techniques. It has been already demonstrated that mRNA levels largely reflect the respective protein levels (32). Consequently, the intersection between proteomic data sets and other genome-wide data sets often allows robust cross-validation. We used independent gene expression data from public datasets to corroborate our findings. It is remarkable that 74% of the differences observed in our proteomics data were confirmed using gene expression data, and are also consistent with previous reports.

MicroRNA expression analysis

Regarding microRNA differential expression analysis, miR-106a, miR-18a and miR-135b expression were higher in TNBC samples, whereas miR-190b, miR-375, miR-205, miR-342, miR-125b and miR-214 were more expressed in ER+ samples. This is consistent with previously published data. miR-342 expression has been beforehand described as significantly higher in ER+ samples, and was

included in a predictor of ER status together with miR-135b and miR-190 (33). Moreover, miR-342 expression was positively correlated with ER α mRNA expression in human breast cancer (34). MiR-205 was significantly under-expressed in triple-negative primary breast cancers when compared with tumor-associated normal samples. Both miR-205 and miR-342 have been proposed as potential biomarkers for diagnosis of TNBC (35). It has been previously showed that miR125b expression is lower in TNBC tumor tissue when compared with adjacent normal tissue, and has been related with chemoresistance (36). Finally, miR-345 expression was down-regulated in ER-negative breast cancer patients, whereas miR-18a was increased (37).

Systems Biology of breast cancers

In this work, we have defined functional nodes and measured their activity, and also performed a comprehensive investigation of miRNA-mediated regulation through an integrative analysis of the variation of miRNA and protein expression in a large collection of breast cancer samples using a probabilistic graphical model approach. The relation among multiple types of molecules defines how they are regulated. Computationally linking and then analyzing these relations at the global scale could reveal entirely new levels of cell regulation and provide insights into damaged molecular pathways in disease (38). Integrated pathway analysis is expected to increase the precision and sensitivity of causal interpretations for large sets of observations since no single data source is likely to provide a complete picture by itself (39).

This analysis identified twelve functional nodes, seven of which showed differential activation after a class comparison analysis when comparing ER+ and TNBC samples. Orthogonal validation using external gene expression datasets confirmed these differences in six out of seven nodes: ECM, proliferation, metabolism A & B and mitochondria A & B. Proliferation and ECM were

grouped in the same branch, despite that activity analyses showed that proliferation activity is increased in TNBC, whereas ECM is decreased. It is well known that TNBCs are highly proliferative tumors with high relapse rate, which is in concordance with our results. Probabilistic graphical model analyses showed that two microRNAs, miR-301a and miR-331-5p, are related with these functional nodes. It has been described that miR-301a promotes breast cancer proliferation, invasion, and tumor growth (40). It has also been described that miR-331 regulates SOCS1 (41), which acts as a negative feedback regulator of STAT (42). ECM node includes a number of proteins related with focal adhesion. There is a previous work showing differences in focal adhesion proteins between ER+ and ER- tumors measured using proteomics (43).

Interestingly, the activity of all metabolism and mitochondria related nodes have a differential activity between ER+ and TNBC samples. Metabolism B and mitochondria B are increased in TNBC samples as compared with ER+ samples, whereas metabolism A and mitochondria A activity are decreased. On the other hand, high expression of LDHB has been associated with glycolytic phenotype (26), whereas it has been demonstrated that loss of FBP1 induces glycolysis and results in increased glucose uptake and macromolecule biosynthesis (44). Cancer cells reprogram their metabolism in order to satisfy their bioenergetics and biosynthetic requirements. The hallmark of tumor metabolism is aerobic glycolysis, or Warburg Effect, which was first described more than 80 years ago (45). During recent years, interest in developing cancer metabolism inhibitors and the possibility of combining them with existing therapeutic approaches has increased dramatically, with academic and pharmaceutical groups actively pursuing this aspect of tumor physiology that could have profound anticancer effects and fundamentally change the treatment of cancer (46). All these results together (glycolysis and the TCA cycle showing differential activity, higher LDHB expression and lower FBP1 expression in TNBC), suggest that TNBC samples are highly glycolytic tumors, and confirm that TNBC have unique metabolic profiles that may be exploited for

therapeutic intervention (47). To support this hypothesis, we evaluated this differential metabolic profile in five breast cancer cell lines. TNBC cell lines showed higher glucose consumption and lactate production, as has been showed previously (26, 44, 47).

Metabolism A node includes some proteins related with glutamine metabolism. Increased glutaminolysis has been described as a key feature of the metabolic changes in cancer cells compared to normal cells (48). Recent studies suggest differences in the metabolism of glutamine between ER- and ER + disease. It has been observed that ER- tumors have a shift in the equilibrium between glutamine and glutamate towards glutamate in the ER- disease (49, 50). We found that GLUD1 expression is slightly lower in TNBC tumors, which is consistent with previous reports describing that TNBCs have the lowest GLUD1 protein expression rate of all molecular subtypes, whereas HER2 positive tumors show the highest (51). We also detected a higher protein expression of P5CS in TNBC tumors. P5CS is a mitochondrial enzyme that catalyzes the conversion of L-glutamate to P5C, a pivotal step in the biosynthesis of non-essential amino acids such as L-proline, L-ornithine and L-arginine (52).

While aerobic proliferating cells use glucose and glutamine for biomass production through the TCA cycle, hypoxic cells shunt glucose to lactate and rewire glutamine metabolism (53). GLUD1 and other glutamate-consuming enzymes activity is stimulated under glucose deprivation conditions, while its activity is repressed during robust glycolysis (54). Glutamine can be used to drive the TCA cycle independently of glucose, contribute to lipid synthesis via IDH1, and be donor for the synthesis of non-essential amino acids -via transaminases and P5CS- and nucleotides (53, 55). Further exploration of glutamine metabolic rewiring differences in subgroups of breast cancer is needed to validate our preliminary results.

It is remarkable that only one protein included in the Metabolism B node (UPGA) was included among the 100 proteins identified as differentially expressed between ER+ and TNBC. This result suggests that functional approaches can highlight differences in biological processes and pathways that cannot be unraveled through the analysis of individual genes or proteins. In this regard, both approaches complement each other.

One of our main objectives was to find relations between microRNA and differential processes in two breast cancer subtypes. We observed that miR-449a expression was lower in TNBC samples, and our probabilistic graphical model analysis include it in the metabolism B node, showing a negative correlation with its activity, which is increased in TNBC samples. Surprisingly, no correlation between miR-449a expression and individual proteins from the node was found. Our results suggest that miR-449a absence is related with and increased glycolytic metabolism, and that miR449a exerts its regulatory effect over cellular metabolism indirectly, through regulation of yet unknown genes (Supplementary Figure 2). Low expression level of miR-449a was correlated with poor prognosis of lung cancer patients (56), but there are no studies about its role in breast cancer progression or its implication in the regulation of the metabolism. miR-449a deserves additional functional studies to investigate its relationship with metabolic rewiring in breast cancers, which is far beyond the scope of this work.

High-throughput technologies are providing a comprehensive view of the molecular changes in cancer tissues. MS-based proteomics together with microRNA profiling allowed us to overview differential processes in ER+ and TNBC samples. However, there are still a number of improvements required before mass spectrometry based proteomics could be considered at the same level than genomics related technologies in terms of comprehensiveness and depth:

substantial advances in speed, sensitivity, reproducibility, better identification methods and standardization of protocols are still necessary.

In conclusion, the integration of different levels of biological information can improve our knowledge of molecular processes in cancer cells. In this study we measured the expression of 1,095 proteins and 90 miRNAs in breast cancer samples, unraveling differences in the activity of molecular processes, such as proliferation and metabolism, between ER+ and TNBC tumors that were also confirmed at the transcriptome level. Moreover, some miRNAs were shown to be related with these functional categories. These findings may provide a rational to design new treatments for TNBC.

Acknowledgements

We want to particularly acknowledge the patients in this study for their participation and to IdiPAZ and I+12 Biobanks for the generous gifts of clinical samples used in this work. We want to thank Dr. Rubén Gómez Rioja for his help with glucose and lactate measurements.

References

1. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer*. 2013;49:1374-403.
2. Espinosa E, Gamez-Pozo A, Sanchez-Navarro I, Pinto A, Castaneda CA, Ciruelos E, et al. The present and future of gene profiling in breast cancer. *Cancer Metastasis Rev*. 2012;31:41-6.
3. Mann M, Kelleher NL. Precision proteomics: the case for high resolution and high mass accuracy. *Proc Natl Acad Sci U S A*. 2008;105:18132-8.
4. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature*. 2012;486:400-4.
5. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. *Nat Rev Genet*. 2012;13:358-69.
6. Gamez-Pozo A, Sanchez-Navarro I, Calvo E, Diaz E, Miguel-Martin M, Lopez R, et al. Protein phosphorylation analysis in archival clinical cancer samples by shotgun and targeted proteomics approaches. *Mol Biosyst*. 2011;7:2368-74.
7. Sprung RW, Jr., Brock JW, Tanksley JP, Li M, Washington MK, Slebos RJ, et al. Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. *Mol Cell Proteomics*. 2009;8:1988-98.
8. Hall JS, Taylor J, Valentine HR, Irlam JJ, Eustace A, Hoskin PJ, et al. Enhanced stability of microRNA expression facilitates classification of FFPE tumour samples exhibiting near total mRNA degradation. *Br J Cancer*. 2012;107:684-94.

9. Liu S, Im H, Bairoch A, Cristofanilli M, Chen R, Deutsch EW, et al. A chromosome-centric human proteome project (C-HPP) to characterize the sets of proteins encoded in chromosome 17. *Journal of proteome research*. 2013;12:45-57.
10. Gamez-Pozo A, Ferrer NI, Ciruelos E, Lopez-Vacas R, Martinez FG, Espinosa E, et al. Shotgun proteomics of archival triple-negative breast cancer samples. *Proteomics Clin Appl*. 2013;7:283-91.
11. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8:118-27.
12. Sanchez-Navarro I, Gamez-Pozo A, Gonzalez-Baron M, Pinto-Marin A, Hardisson D, Lopez R, et al. Comparison of gene expression profiling by reverse transcription quantitative PCR between fresh frozen and formalin-fixed, paraffin-embedded breast cancer tissues. *Biotechniques*. 2010;48:389-97.
13. Dweep H, Sticht C, Pandey P, Gretz N. miRWalk--database: prediction of possible miRNA binding sites by "walking" the genes of three genomes. *J Biomed Inform*. 2011;44:839-47.
14. Schwarz G. Estimating the dimension of a model. *Annals of Statistics*. 1978;6:461-4.
15. Lauritzen S. *Graphical Models*. Oxford, U.K.: Oxford University Press.; 1996.
16. Chow CK, Liu CN. Approximating discrete probability distributions with dependence trees. *IEEE Transactions Information Theory*. 1968;14:462-7.
17. Kruskal JB. On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society*. 1956;7:48-50.
18. Cover TM, Thomas JA. *Elements of Information Theory*, 2nd ed. . New York: Wiley Intersci.; 2006.
19. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria.: R Foundation for Statistical Computing; 2013.

20. Abreu GCG, Edwards D, Labouriau R. High-Dimensional Graphical Model Search with the gRapHD R Package. *Journal of Statistical Software*. 2010;37:1-18.
21. Miller LD, Coffman LG, Chou JW, Black MA, Bergh J, D'Agostino R, Jr., et al. An iron regulatory gene signature predicts outcome in breast cancer. *Cancer Res*. 2011;71:6728-37.
22. Guedj M, Marisa L, de Reynies A, Orsetti B, Schiappa R, Bibeau F, et al. A refined molecular taxonomy of breast cancer. *Oncogene*. 2012;31:1196-206.
23. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490:61-70.
24. Wang M, Weiss M, Simonovic M, Haertinger G, Schrimpf SP, Hengartner MO, et al. PaxDb, a database of protein abundance averages across all three domains of life. *Mol Cell Proteomics*. 2012;11:492-500.
25. Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature*. 2011;476:346-50.
26. Dennison JB, Molina JR, Mitra S, Gonzalez-Angulo AM, Balko JM, Kuba MG, et al. Lactate dehydrogenase B: a metabolic marker of response to neoadjuvant chemotherapy in breast cancer. *Clin Cancer Res*. 2013;19:3703-13.
27. McKiernan E, McDermott EW, Evoy D, Crown J, Duffy MJ. The role of S100 genes in breast cancer progression. *Tumour Biol*. 2011;32:441-50.
28. Cook KL, Shajahan AN, Warri A, Jin L, Hilakivi-Clarke LA, Clarke R. Glucose-regulated protein 78 controls cross-talk between apoptosis and autophagy to determine antiestrogen responsiveness. *Cancer Res*. 2012;72:3337-49.
29. Fletcher GC, Patel S, Tyson K, Adam PJ, Schenker M, Loader JA, et al. hAG-2 and hAG-3, human homologues of genes involved in differentiation, are associated with oestrogen receptor-

positive breast tumours and interact with metastasis gene C4.4a and dystroglycan. *Br J Cancer*. 2003;88:579-85.

30. Baniwal SK, Ching NO, Jordan VC, Tripathy D, Frenkel B. Prolactin-induced protein (PIP) regulates proliferation of luminal A type breast cancer cells in an estrogen-independent manner. *PLoS One*. 2013;8:e62361.

31. Descotes F, Jezequel P, Spyrtos F, Campion L, Grenot C, Lerebours F, et al. Identification of potential prognostic biomarkers for node-negative breast tumours by proteomic analysis: a multicentric 2004 national PHRC study. *Int J Oncol*. 2012;41:92-104.

32. Marguerat S, Schmidt A, Codlin S, Chen W, Aebersold R, Bahler J. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell*. 2012;151:671-83.

33. Lowery AJ, Miller N, Devaney A, McNeill RE, Davoren PA, Lemetre C, et al. MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. *Breast Cancer Res*. 2009;11:R27.

34. He YJ, Wu JZ, Ji MH, Ma T, Qiao EQ, Ma R, et al. miR-342 is associated with estrogen receptor- α expression and response to tamoxifen in breast cancer. *Exp Ther Med*. 2013;5:813-8.

35. Savad S, Mehdipour P, Miryounesi M, Shirkoobi R, Fereidooni F, Mansouri F, et al. Expression analysis of MiR-21, MiR-205, and MiR-342 in breast cancer in Iran. *Asian Pac J Cancer Prev*. 2012;13:873-7.

36. Ouyang M, Li Y, Ye S, Ma J, Lu L, Lv W, et al. MicroRNA profiling implies new markers of chemoresistance of triple-negative breast cancer. *PLoS One*. 2014;9:e96228.

37. Jonsdottir K, Janssen SR, Da Rosa FC, Gudlaugsson E, Skaland I, Baak JP, et al. Validation of expression patterns for nine miRNAs in 204 lymph-node negative breast cancers. *PLoS One*. 2012;7:e48692.
38. Wiley HS. Integrating multiple types of data for signaling research: challenges and opportunities. *Sci Signal*. 2011;4:pe9.
39. Vaske CJ, Benz SC, Sanborn JZ, Earl D, Szeto C, Zhu J, et al. Inference of patient-specific pathway activities from multi-dimensional cancer genomics data using PARADIGM. *Bioinformatics*. 2010;26:i237-45.
40. Shi W, Gerster K, Alajez NM, Tsang J, Waldron L, Pintilie M, et al. MicroRNA-301 mediates proliferation and invasion in human breast cancer. *Cancer Res*. 2011;71:2926-37.
41. Zanette DL, Rivadavia F, Molfetta GA, Barbuzano FG, Proto-Siqueira R, Silva-Jr WA, et al. miRNA expression profiles in chronic lymphocytic and acute lymphocytic leukemia. *Braz J Med Biol Res*. 2007;40:1435-40.
42. Rakesh K, Agrawal DK. Cytokines and growth factors involved in apoptosis and proliferation of vascular smooth muscle cells. *Int Immunopharmacol*. 2005;5:1487-506.
43. Liu NQ, Dekker LJ, Stingl C, Guzel C, De Marchi T, Martens JW, et al. Quantitative proteomic analysis of microdissected breast cancer tissues: comparison of label-free and SILAC-based quantification with shotgun, directed, and targeted MS approaches. *J Proteome Res*. 2013;12:4627-41.
44. Dong C, Yuan T, Wu Y, Wang Y, Fan TW, Miriyala S, et al. Loss of FBP1 by Snail-mediated repression provides metabolic advantages in basal-like breast cancer. *Cancer Cell*. 2013;23:316-31.
45. Warburg O. On the origin of cancer cells. *Science*. 1956;123:309-14.
46. Jones NP, Schulze A. Targeting cancer metabolism--aiming at a tumour's sweet-spot. *Drug Discov Today*. 2012;17:232-41.

47. McClelland ML, Adler AS, Shang Y, Hunsaker T, Truong T, Peterson D, et al. An integrated genomic screen identifies LDHB as an essential gene for triple-negative breast cancer. *Cancer Res.* 2012;72:5812-23.
48. Lu W, Pelicano H, Huang P. Cancer metabolism: is glutamine sweeter than glucose? *Cancer Cell.* 2010;18:199-200.
49. Budczies J, Pfitzner BM, Gyorffy B, Winzer KJ, Radke C, Dietel M, et al. Glutamate enrichment as new diagnostic opportunity in breast cancer. *Int J Cancer.* 2014.
50. Cao MD, Lamichhane S, Lundgren S, Bofin A, Fjosne H, Giskeodegard GF, et al. Metabolic characterization of triple negative breast cancer. *BMC Cancer.* 2014;14:941.
51. Kim S, Kim do H, Jung WH, Koo JS. Expression of glutamine metabolism-related proteins according to molecular subtype of breast cancer. *Endocr Relat Cancer.* 2013;20:339-48.
52. Hu CA, Khalil S, Zhaorigetu S, Liu Z, Tyler M, Wan G, et al. Human Delta1-pyrroline-5-carboxylate synthase: function and regulation. *Amino Acids.* 2008;35:665-72.
53. Dang CV. Links between metabolism and cancer. *Genes Dev.* 2012;26:877-90.
54. Yang C, Sudderth J, Dang T, Bachoo RM, McDonald JG, DeBerardinis RJ. Glioblastoma cells require glutamate dehydrogenase to survive impairments of glucose metabolism or Akt signaling. *Cancer Res.* 2009;69:7986-93.
55. DeBerardinis RJ, Cheng T. Q's next: the diverse functions of glutamine in metabolism, cell biology and cancer. *Oncogene.* 2010;29:313-24.
56. Luo W, Huang B, Li Z, Li H, Sun L, Zhang Q, et al. MicroRNA-449a is downregulated in non-small cell lung cancer and inhibits migration and invasion by targeting c-Met. *PLoS One.* 2013;8:e64759.

Figures Legends

Figure 1: Hierarchical cluster of proteins differentially expressed between ER+ (orange) and TNBC (pink) samples with an FDR<2% identified by Significance Analysis of Microarrays. Red means high expression and green low expression. Two main clusters of proteins can be observed, one up-regulated (above) and other down-regulated (below) in TNBCs. Protein position in the cluster can be found in supplementary table 3.

Figure 2: Hierarchical cluster of microRNAs differentially expressed between ER+ (orange) and TNBC (pink) samples with an FDR<5% identified by Significance Analysis of Microarrays. Red means high expression and green means low expression. Two main clusters of microRNAs can be observed, one up-regulated (above) and other down-regulated (below) in TNBCs.

Figure 3: Probabilistic graphical model analysis unravel the functional organization of proteins and microRNAs in breast cancers. Branches with no functional structure and the core remain in gray, whereas colored branches harbor one or two functional nodes. Big gray circles correspond to microRNAs.

Figure 4: Probabilistic graphical model showing median expression values of proteins and microRNAs in ER+ (A) and TNBC (B) samples. Red means high expression and green means low expression. The magnified area corresponds to the metabolism B node in ER+ (C) and TNBC (D). Big circles correspond to microRNAs.

Figure 5: Metabolic activity of ER+ (white) and TNBC (gray) cell lines. A) Glucose consumption is measured in mg/dL. Kruskal-Wallis test indicates that differences between ER+ and TNBC cell lines are significant ($p<0.05$). B) Lactate production is measured in mmol/L. Kruskal-Wallis test indicates that differences between ER+ and TNBC cell lines are significant ($p<0.01$).

Table 1. Clinical characteristics of the patients included in the study. Clinical criteria are provided according to TNM classification.

(<http://www.cancer.gov/cancertopics/pdq/treatment/breast/healthprofessional/page3>). Tumor grade is the description of a tumor based on how abnormal the tumor cells and the tumor tissue look under a microscope, for determination procedure, see <http://www.cancer.gov/cancertopics/factsheet/detection/tumor-grade>.

	All	ER+	TNBC
Age at diagnosis (median)	54.6 (32-83)	54.2 (32-83)	61.2 (37-78)
Age at diagnosis (mean)	55.2	54.1	58.5
Tumor Size			
T1	33 (31%)	28 (35%)	5 (19%)
T2	61 (58%)	42 (53%)	19 (73%)
T3	10 (9%)	8 (10%)	2 (8%)
T4	1 (1%)	1 (1%)	0 (0%)
Multifocal	1 (1%)	1 (1%)	0 (0%)
Tumor Grade			
G1	12 (11%)	12 (15%)	0 (0%)
G2	33 (31%)	29 (36%)	4 (15%)
G3	41 (39%)	21 (26%)	20 (77%)
Unknown	20 (19%)	18 (23%)	2 (8%)
Lymph node status			
N0	0 (0%)	0 (0%)	0 (0%)
N1	71 (67%)	54 (68%)	17 (65%)
N2	35 (33%)	26 (32%)	9 (35%)
Chemotherapy			
No Anthraciclynes	34 (32%)	23 (29%)	11 (42%)
Anthraciclynes	63 (59%)	51 (64%)	12 (46%)
Anthraciclynes + taxanes	9 (9%)	6 (7%)	3 (12%)

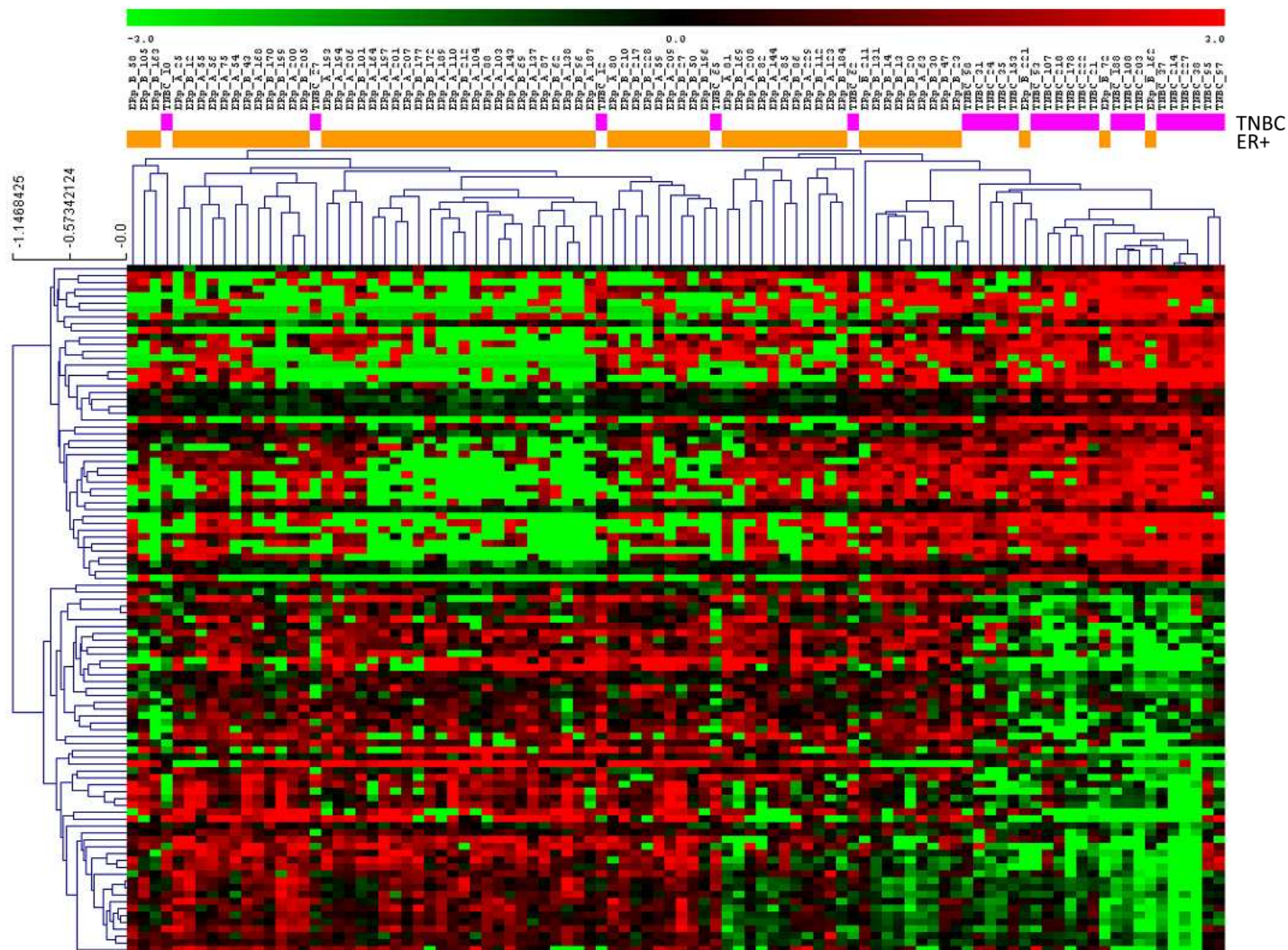


Figure 1

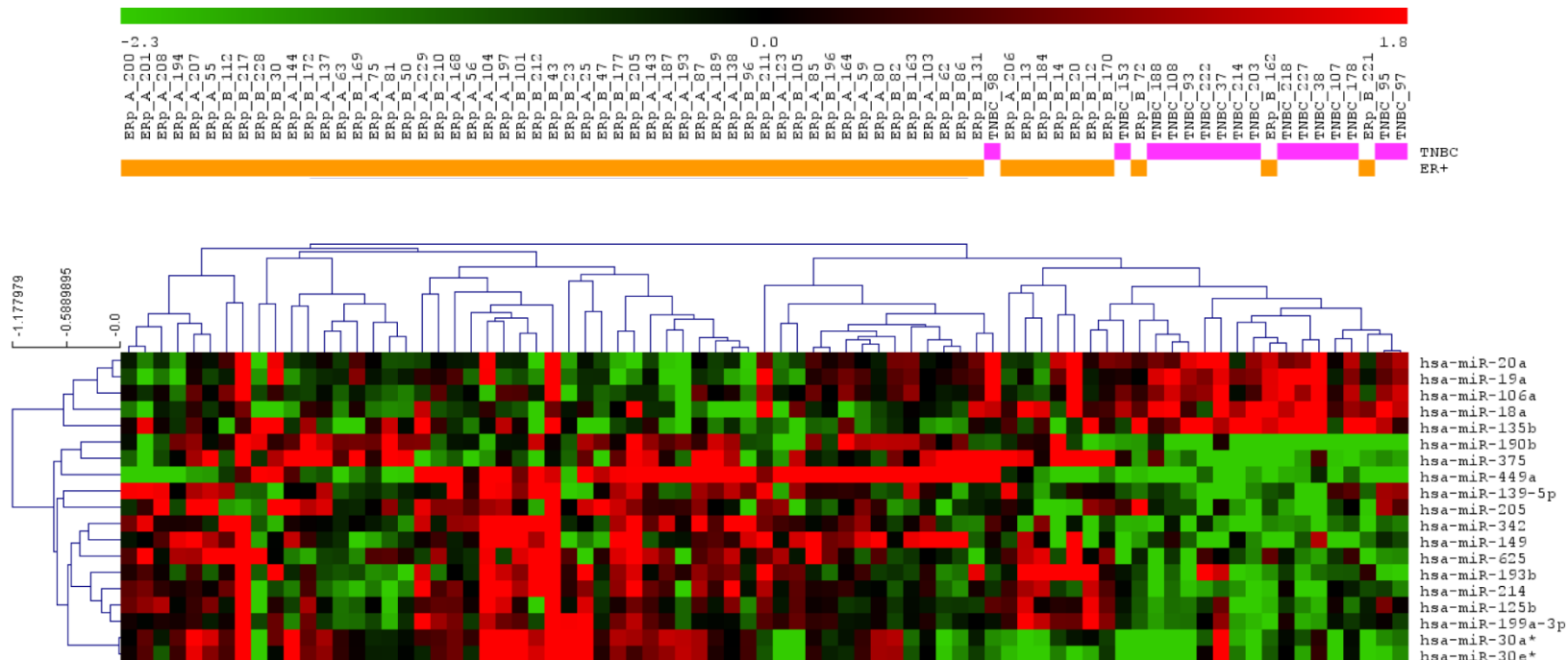


Figure 2

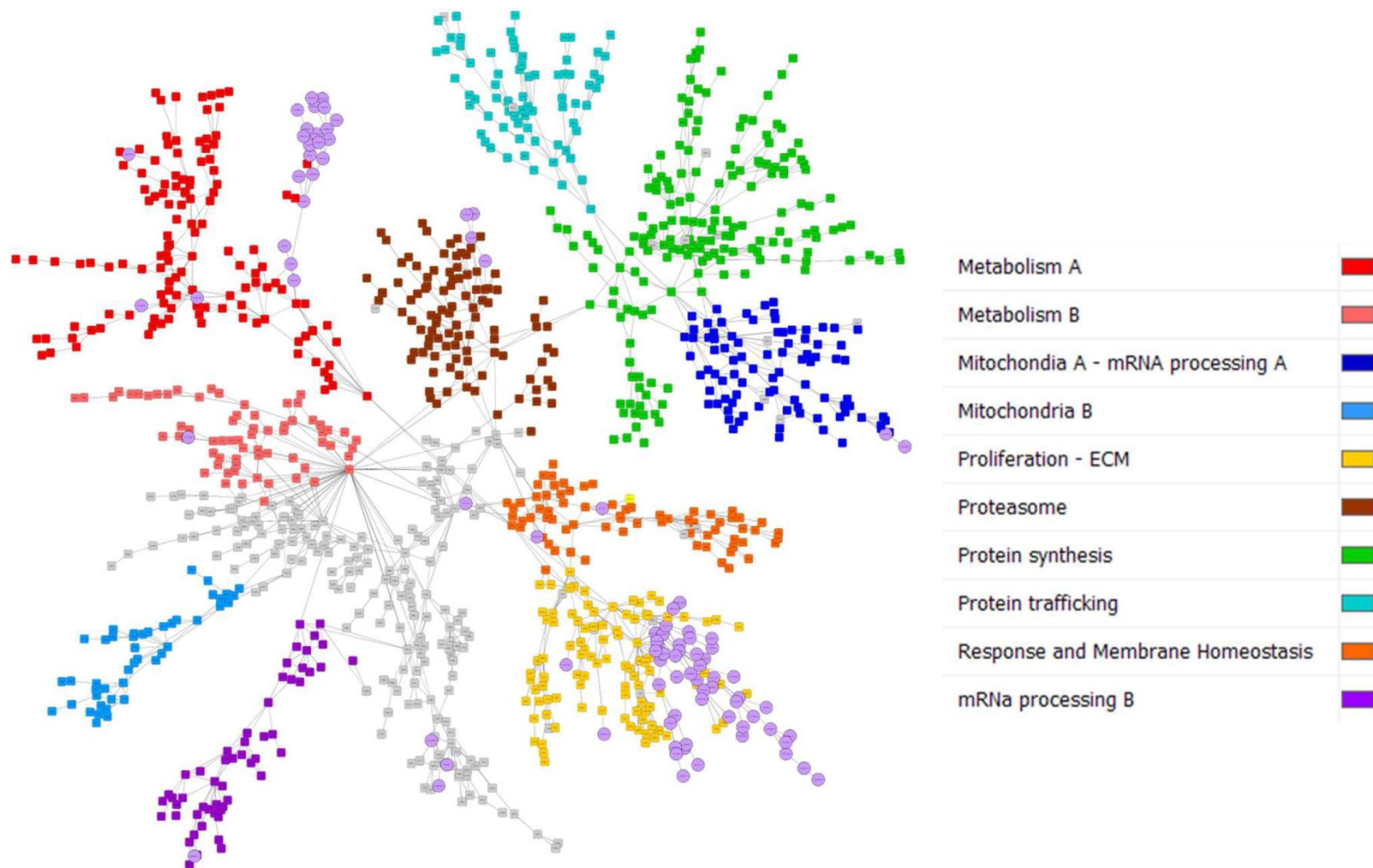


Figure 3

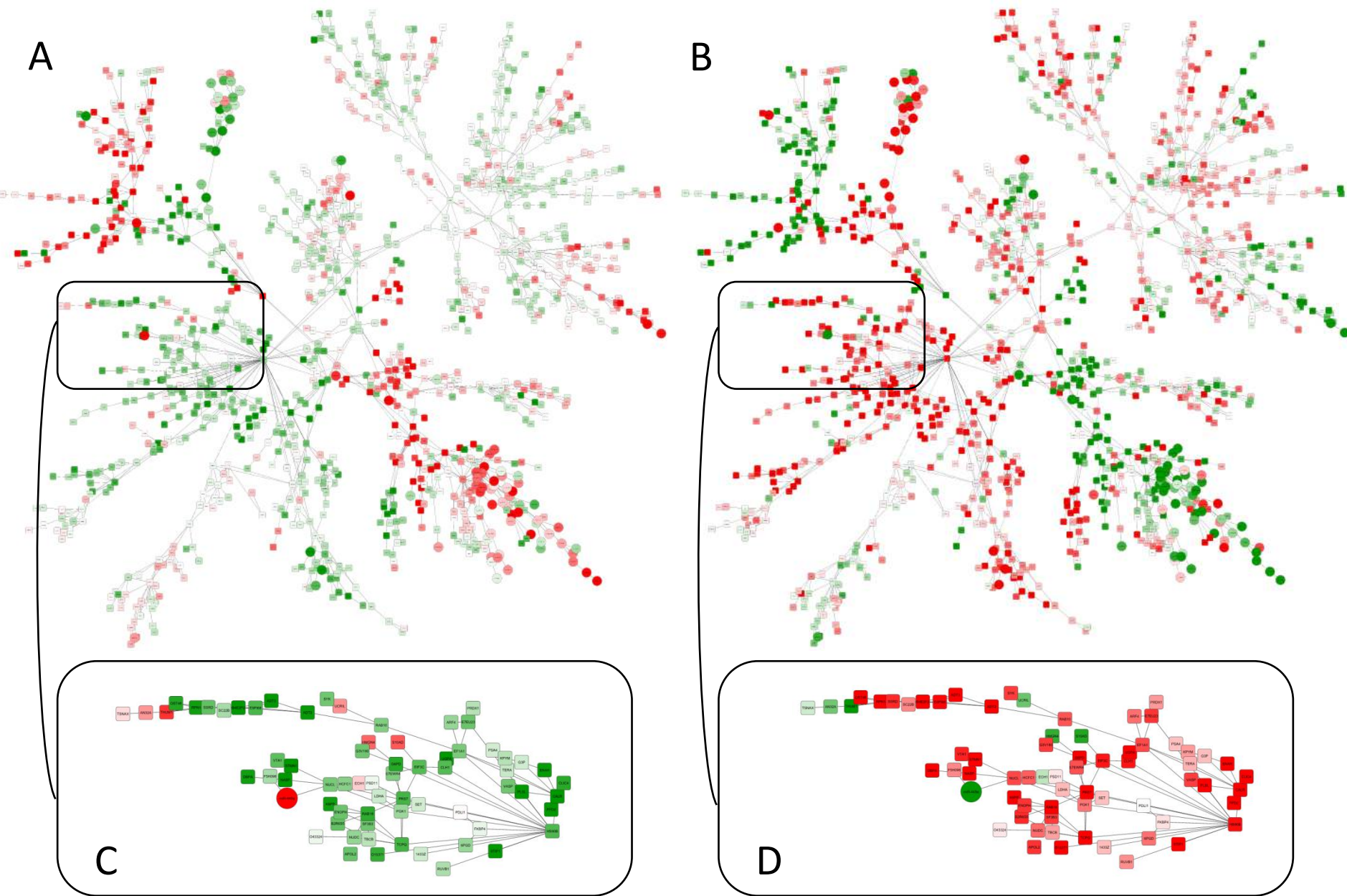
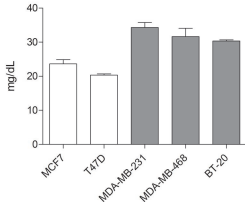
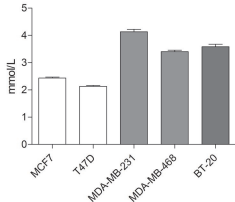


Figure 4

A**Glucose consumption****B****Lactate production****Figure 5**Downloaded from cancerres.aacr.org/

Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

Combined label-free quantitative proteomics and microRNA expression analysis of breast cancer unravel molecular differences with clinical implications.

Angelo Gámez-Pozo, Julia Berges-Soria, Jorge M Arevalillo, et al.

Cancer Res Published OnlineFirst April 16, 2015.

Updated version	Access the most recent version of this article at: doi: 10.1158/0008-5472.CAN-14-1937
Supplementary Material	Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2015/04/17/0008-5472.CAN-14-1937.DC1.html
Author Manuscript	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org .